Live Archive System

Jason Hyon

Jet Propulsion Laboratory,
California Institute of Technology
Jason.hyon@jpl.nasa.gov

Trends

- Space Sciences are confronted with overwhelming volume of data
  - Granularity of registered observations
  - Modeling based on High performance computing
  - Data On-demand via network
- Inexpensive emerging storage technologies, combined with the availability of high-speed communications will offer the infrastructure for extremely large data repositories to be accessible on-line.

Challenge

- Rising volume of data that needs to be safely distributed and archived
- Current technologies of CD and DVD no longer viable solutions
- Magnetic media are not viable
- Higher capacity media must be found
- An alternative should exist

NASA Standards for Media

- The National Archives and Records Administration (NARA) and the National Institute of Standards and Technologies (NIST) determine the suitability of media for archiving.
- They only publish that which has proven to be reliable which is a long-term process.
  - For example, the acceptance of CD-ROM took around 10 years.

A Guideline for Archive Media

- Must use an open standard (UDF/ISO 9660)
- Should have multiple vendors of hardware and media
- Must carry directory information
- Must employ robust error correction with graceful degradation
- Must provide the capability for easy migration to higher capacity media
- Should have some organization monitoring industry performance

Status of Current Archive Technology

- There is no cost-effective archive solution
  - Size is limited and cost is high
  - Data transfer rate is slow
  - It needs special environment

- A concept of “live” media
  - Compatible physical format for next 5 – 10 years(upward compatible)
  - Compatible logical format for next 5 – 10 years (ISO/de facto)
  - Migration path is well defined

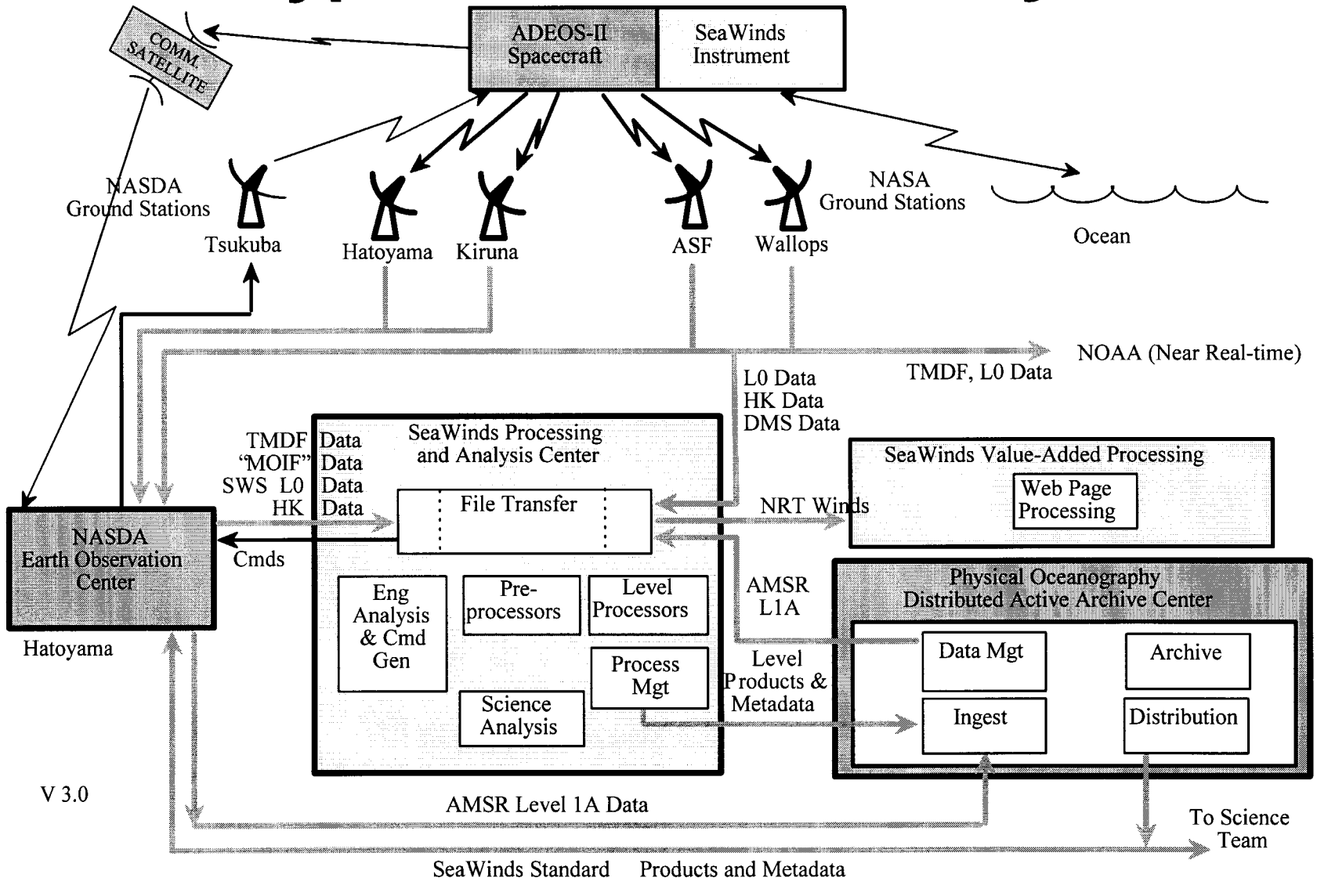
Our Approach

- Factors
  - Archive storage technology tends to lag 5 years compared to real-time storage technology.
  - There tend to be two separate systems for data processing and distribution.
- A combination of the two systems will be cost effective and efficient.
  - State-of-the-art storage technology
  - Longevity

Live Archive System

- We have developed a storage architecture that addresses:
  - Hardware Architecture
  - Software Architecture
  - Operations and Processes
  - Hardware Technology
  - Storage Area Management Software
  - Technology Migration
- Based on Open Archive Information System (OAIS)
  - http://www.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf
- To combine data processing, distribution, and archiving systems for a single efficient data storage for all purpose

A Typical Ground Data System



Hardware Architecture

- Infrastructure to interface hardware components and to modify existing components seamlessly
  - On-line, Near on-line, Off-line, Bankup system, Order system
- Interface Specifications
  - SCSI, FC, TCP/IP, Net Apps, Security
- Must support the OAIS model

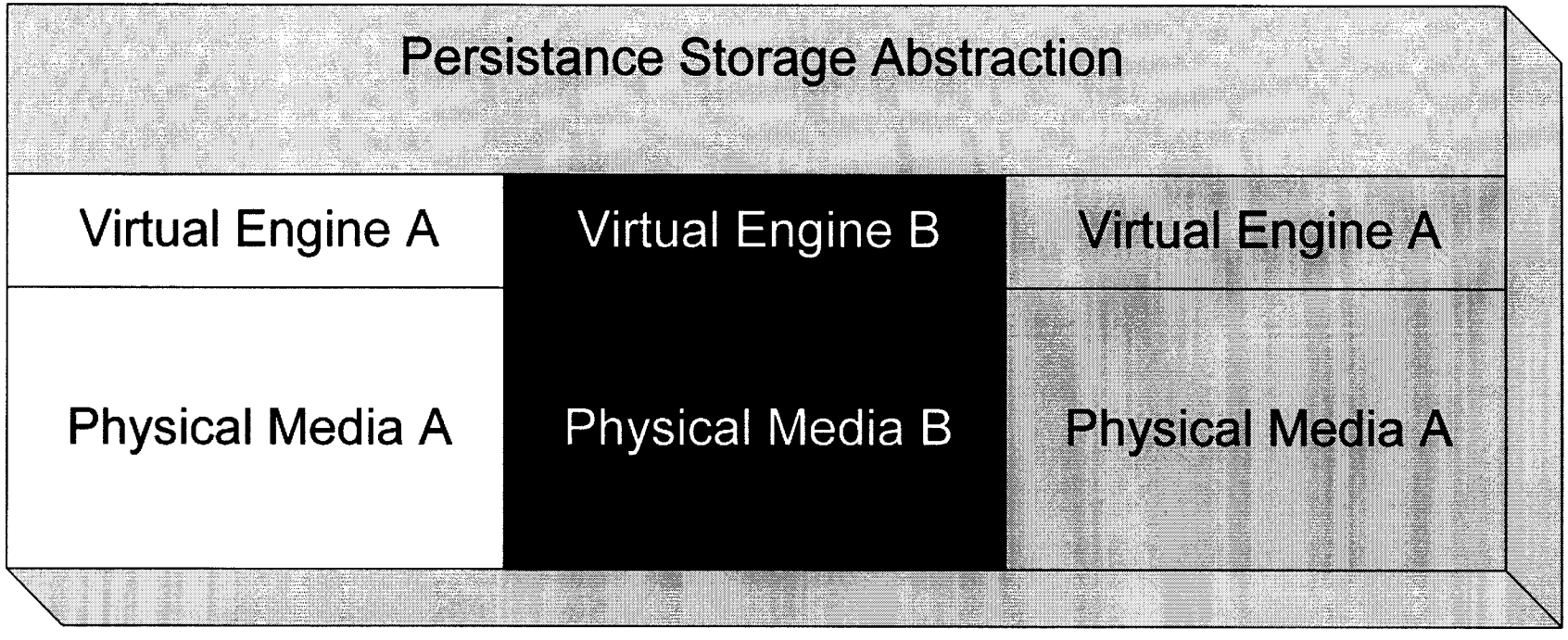


Fig. 5 Physical Storage Abstraction

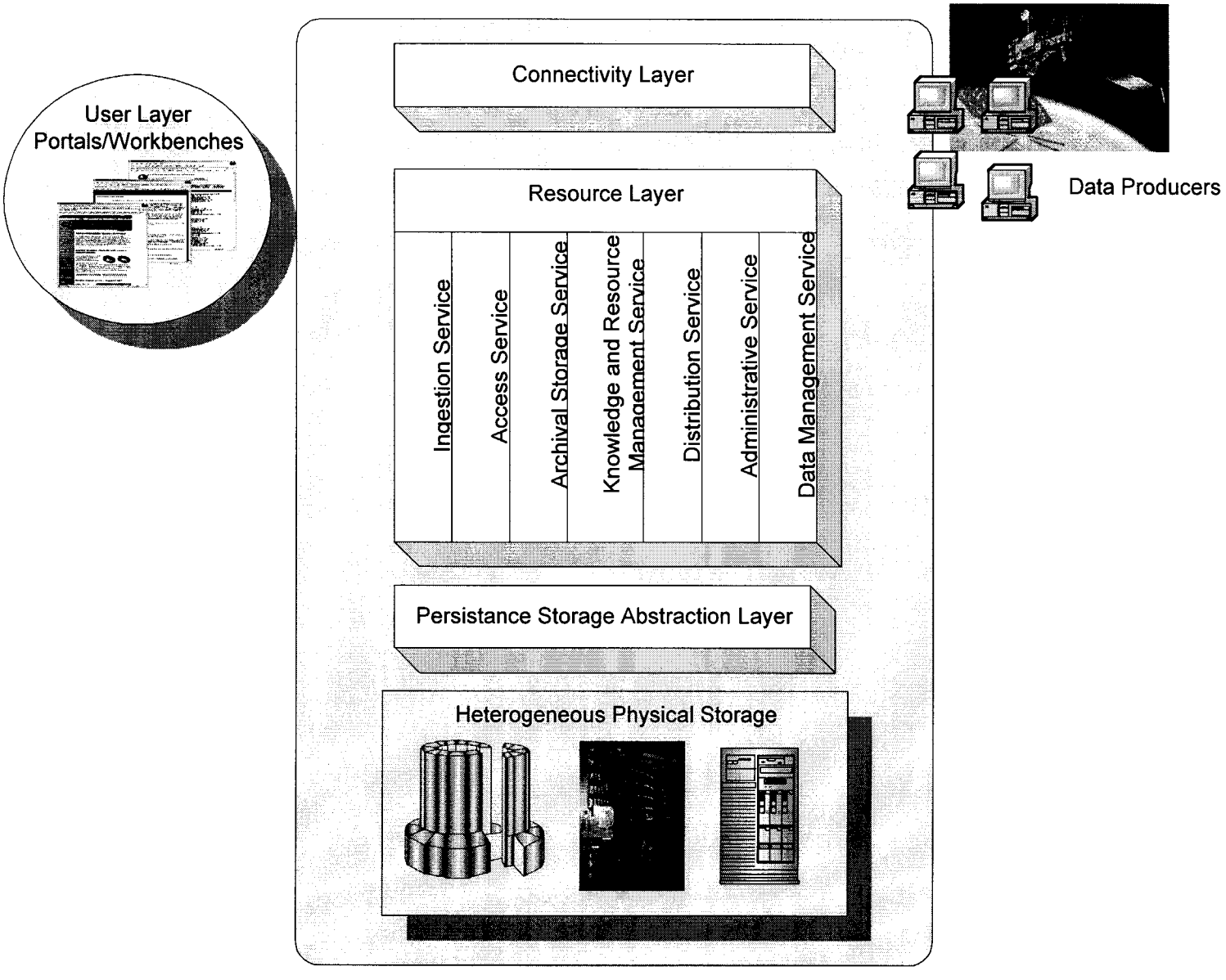
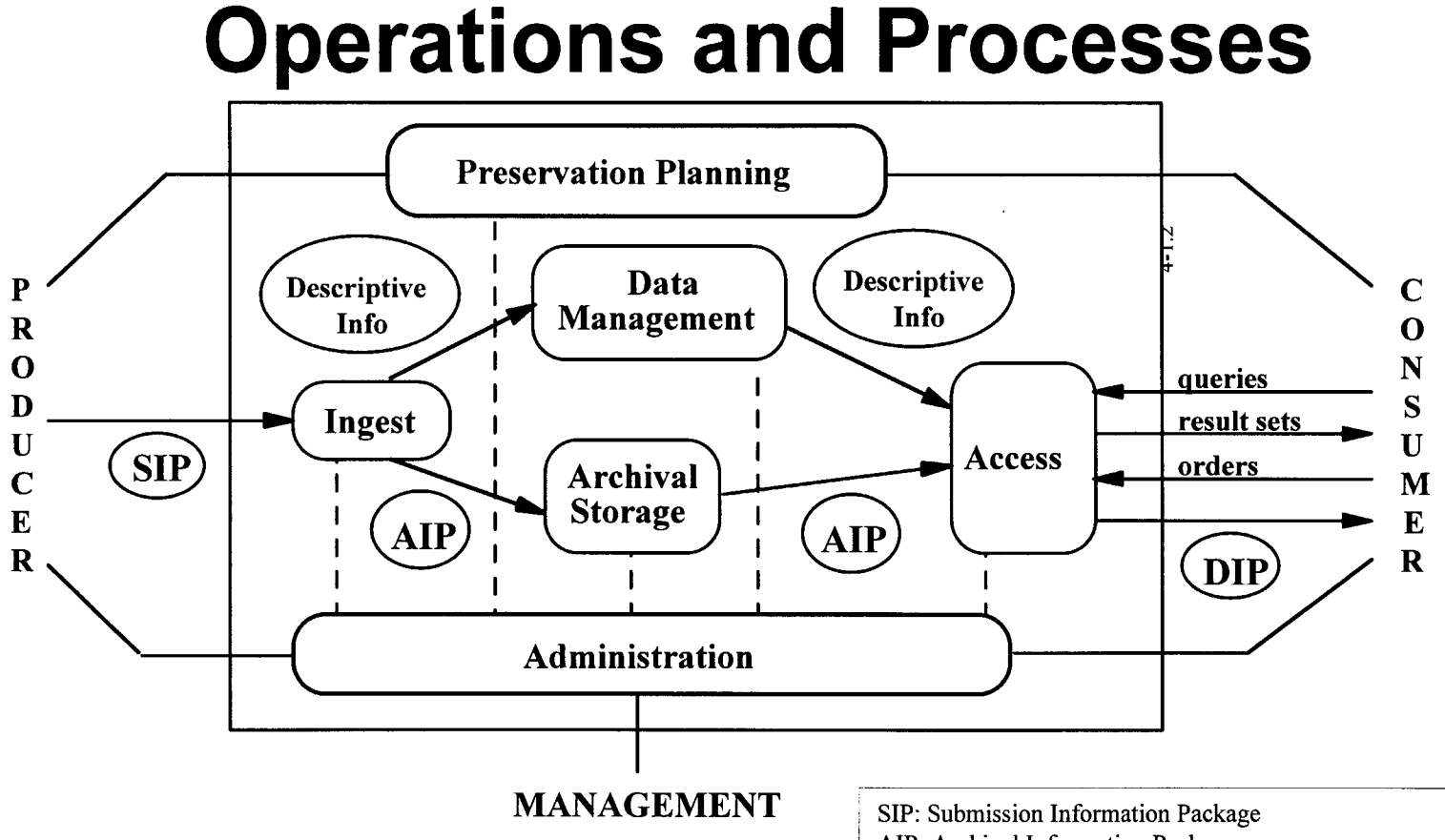


Fig. 6 Live Mission System(LMS) Architecture



SIP: Submission Information Package
AIP: Archival Information Package
DIP: Dissemination Information Package

Fig. 1 OAIS Functional Model

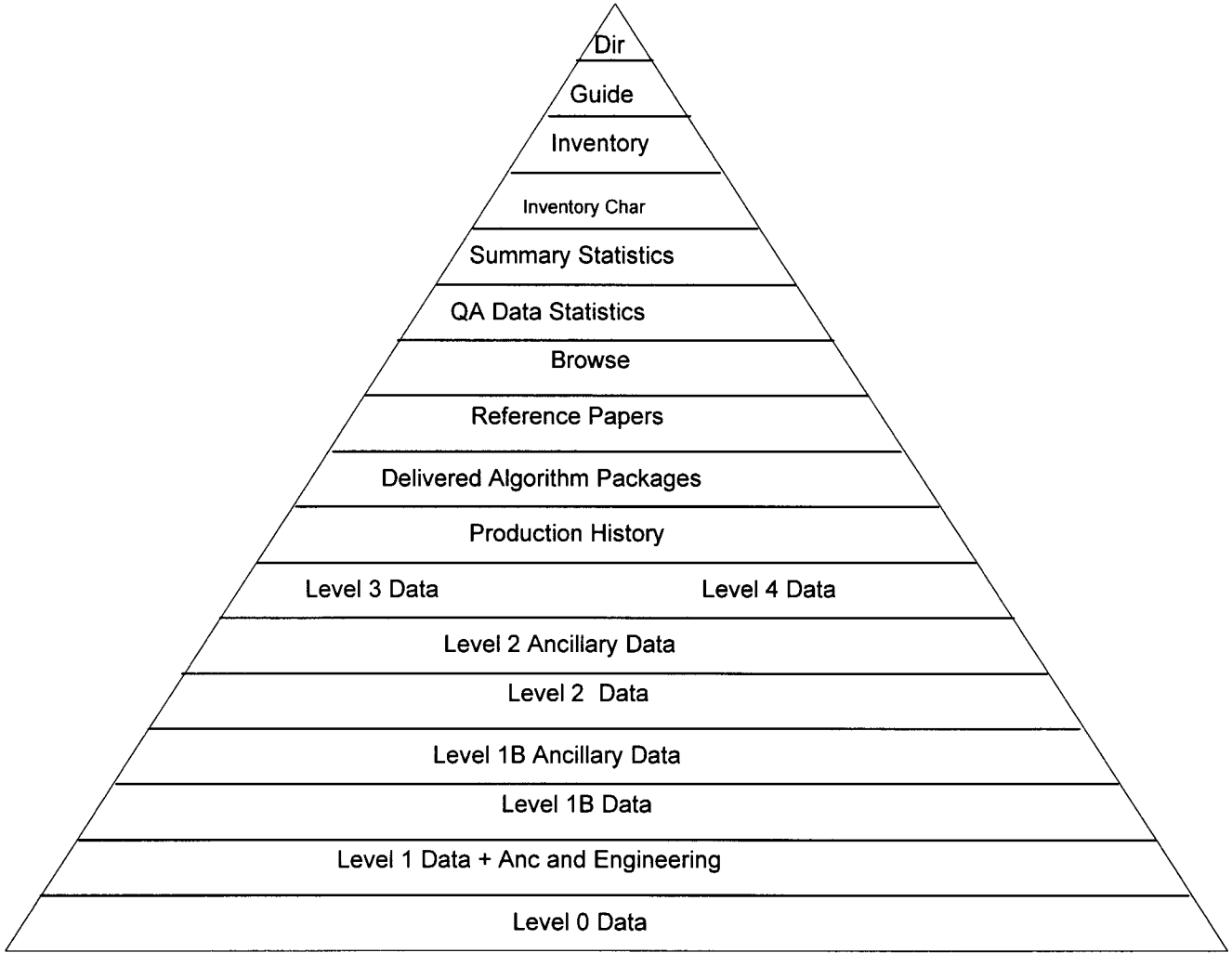


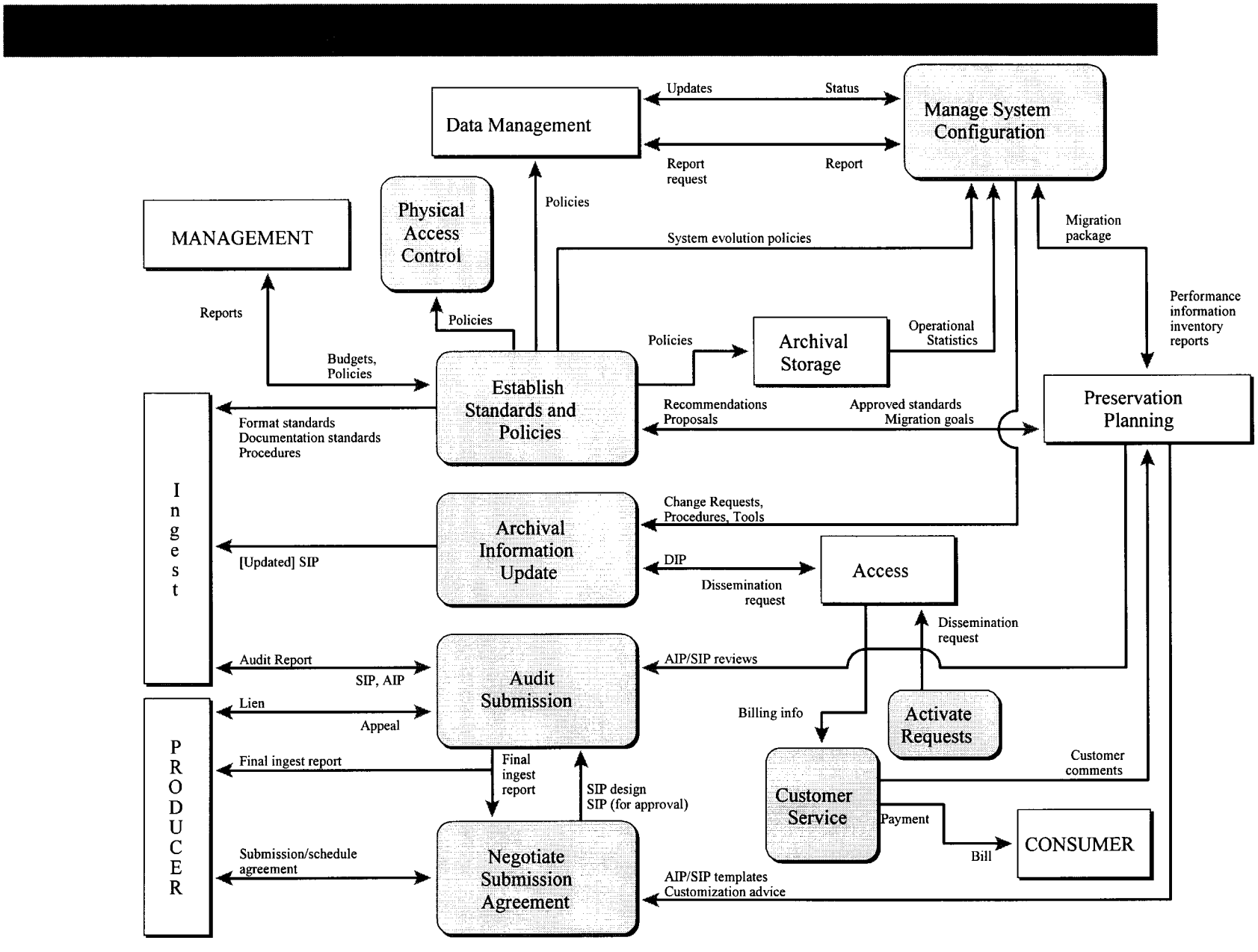
Fig. 4 Science Date Processing Levels

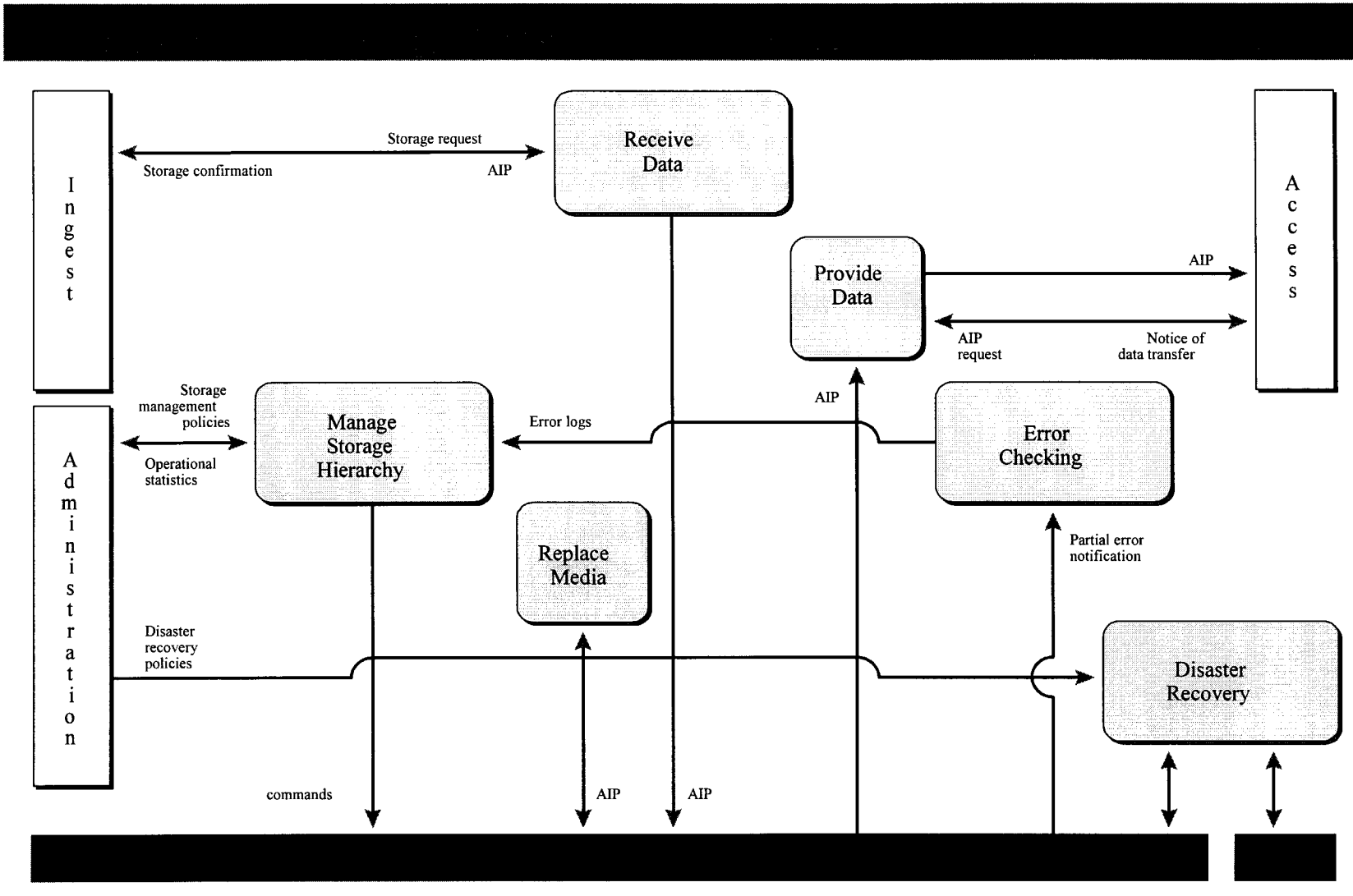
Hardware Technology

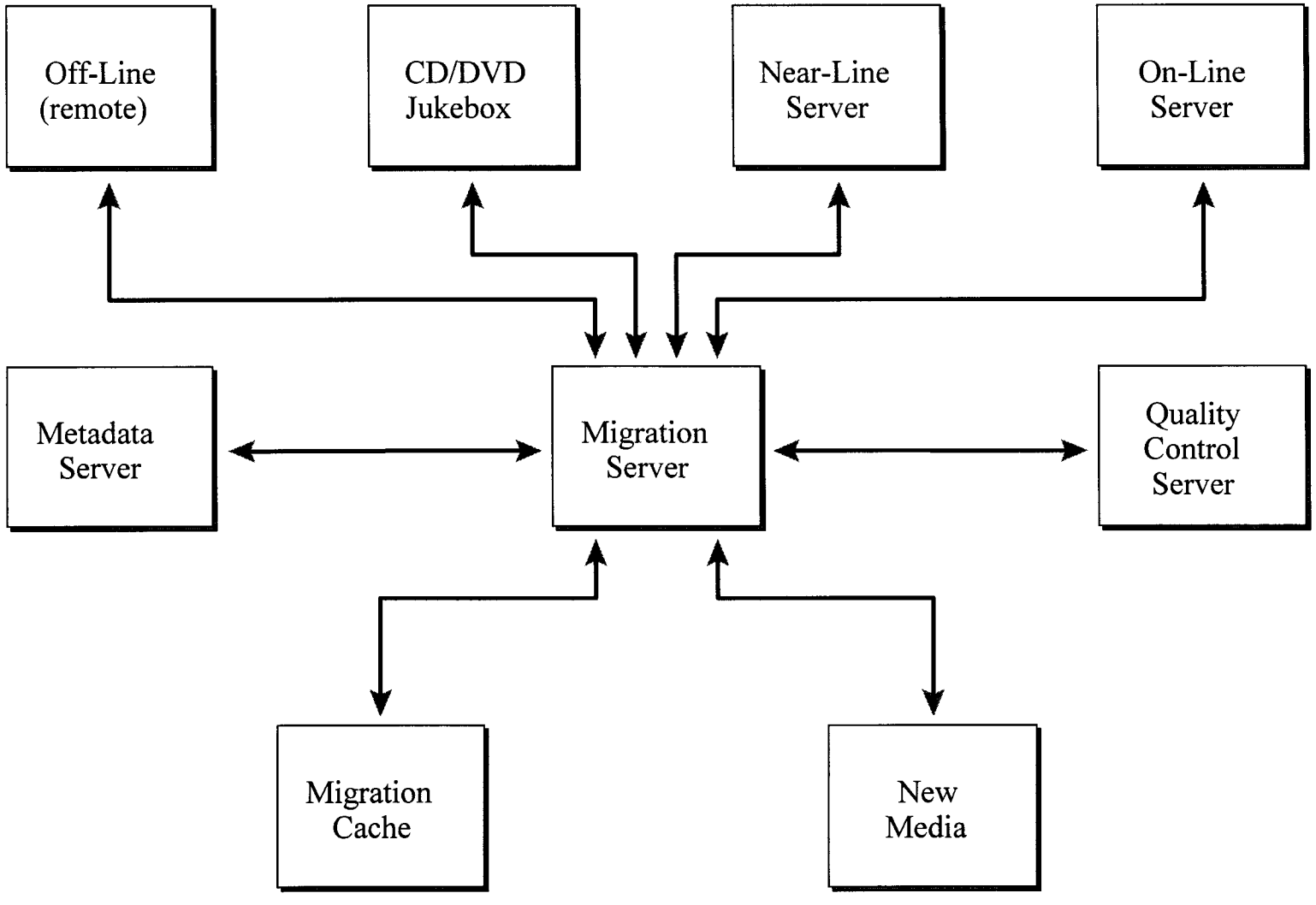
- Synchronous Mirroring: SCSI/FC
- Asynchronous Mirroring or Replication: SCSI/FC
- Snapshot Data Protection: ATA disk
- Backup: Tape
- Archiving: CD/DVD
- Distribution: CD/DVD/Tape/Internet

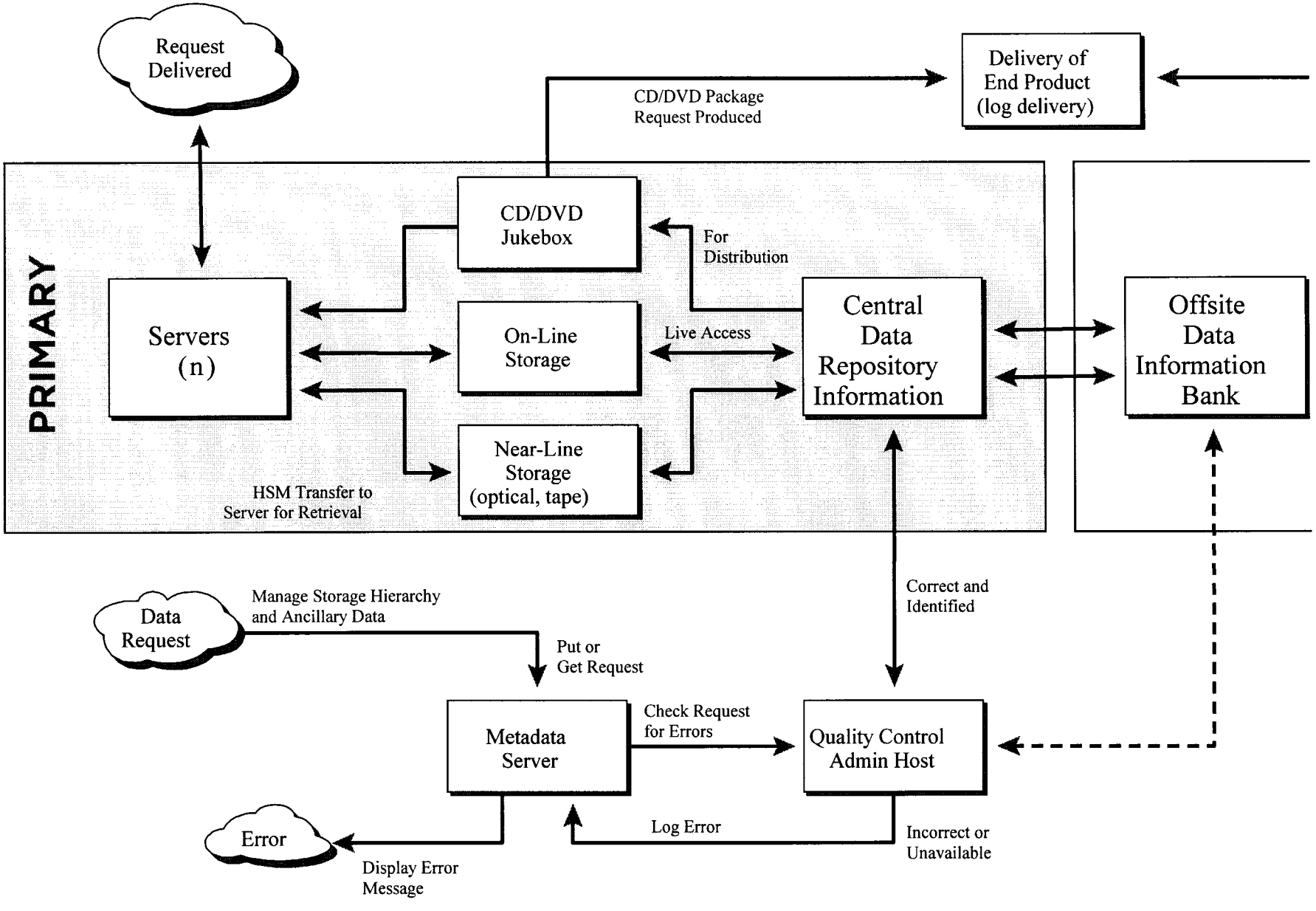
Storage Area Network Management

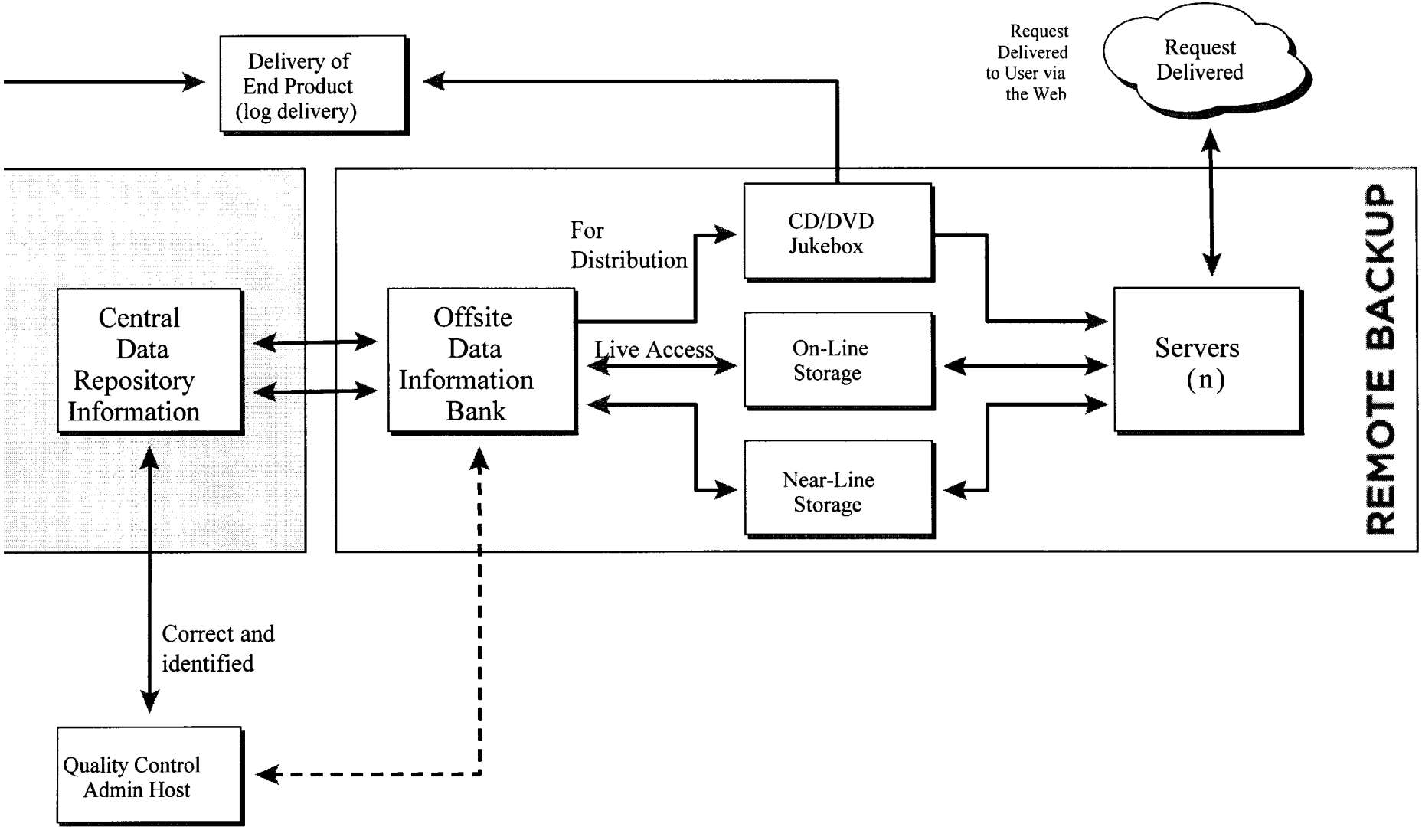
- System Administration
- Realtime and Archival data processing
- Media migration server
- Backup server


Data Management
Updates
Status
Manage System Configuration
Report request
Report
Physical Access Control
Policies
MANAGEMENT
System evolution policies
Migration package
Performance information inventory reports
Reports
Policies
Policies
Archival Storage
Operational Statistics
Budgets, Policies
Establish Standards and Policies
Preservation Planning
Recommendations Proposals
Approved standards Migration goals
Format standards
Documentation standards
Procedures
Ingest
Change Requests, Procedures, Tools
Archival Information Update
[Updated] SIP
DIP
Dissemination request
Access
Dissemination request
Audit Report
SIP, AIP
AIP/SIP reviews
Audit Submission
Lien
Appeal
Billing info
Activate Requests
PRODUCER
Final ingest report
Final ingest report
SIP design
SIP (for approval)
Customer comments
Customer Service
Payment
Bill
CONSUMER
Submission/schedule agreement
Negotiate Submission Agreement
AIP/SIP templates
Customization advice


Ingest
Storage request
Storage confirmation
AIP
Receive Data
Provide Data
AIP
Access
AIP request
Notice of data transfer
Administration
Storage management policies
Operational statistics
Manage Storage Hierarchy
Error logs
AIP
Error Checking
Replace Media
Partial error notification
Disaster recovery policies
Disaster Recovery
commands
AIP
AIP


Off-Line
(remote)
CD/DVD
Jukebox
Near-Line
Server
On-Line
Server
Metadata
Server
Migration
Server
Quality
Control
Server
Migration
Cache
New
Media


Request Delivered
Delivery of End Product (log delivery)
CD/DVD Package Request Produced
PRIMARY
CD/DVD Jukebox
For Distribution
Servers (n)
On-Line Storage
Live Access
Central Data Repository Information
Offsite Data Information Bank
Near-Line Storage (optical, tape)
HSM Transfer to Server for Retrieval
Data Request
Manage Storage Hierarchy and Ancillary Data
Put or Get Request
Correct and Identified
Metadata Server
Check Request for Errors
Quality Control Admin Host
Error
Display Error Message
Log Error
Incorrect or Unavailable


Delivery of
End Product
(log delivery)
Request
Delivered
to User via
the Web
Request
Delivered
REMOTE BACKUP
CD/DVD
Jukebox
For
Distribution
Central
Data
Repository
Information
Offsite
Data
Information
Bank
Live Access
On-Line
Storage
Servers
(n)
Near-Line
Storage
Correct and
identified
Quality Control
Admin Host

Conclusion

- Flexible hardware and software architecture with rigorous system administration process
- OAIS Model

Will provide a combined system to support both real-time needs and archival needs to enable efficient data storage systems.